*ie*Scope

The Web as the World

ieWiki Inc.
11685 Kistner Road, San Diego
San Diego, CA, 92128.
January 3, 2000

Exhibit B

*i*eScope – Mining the Internet

This document describes a new concept, *i*eScope, which mines the Internet to augment data sources with real world information.

Today the Internet is the most up-to-date and complete description of the world. If there is something happening, or you are interested in finding a specific piece of information, more and more people are turning *intuitively* to the Internet. The Internet now contains an incredible array of information, from real-time data such as news, sport, financial and weather to long-term sources of description about the world such as corporate home pages, product information, geographic and demographic descriptions, encyclopedias, dictionaries and other resource material. As of late 1999, the Internet holds about 1 billion pages of information, and this figure is growing dramatically.

The world today is focused on the Internet as a medium for communication, interaction, distribution and transaction. *i*eScope is a philosophical shift – viewing the Internet as the world's largest and most up-to-date description of the world.

Every customer-oriented company is hungry for information on its customers – what better place to search than the Internet. *i*eScope works by looking at each transaction in a company's database, and searching through the Internet to augment the data available on this transaction and its customers. This data augmentation adds not only well-defined descriptive keywords to each transaction, but also keywords that work towards describing the key values behind the transaction.

*i*eScope can be used in a number of different ways, each of which will be discussed in further detail. As a segmentation engine, *i*eScope can search through an internet-enhanced customer database to find people in many different categories, for example, those who have just had children, who may be interested in investing in pharmaceutical stocks, or who enjoy baseball.

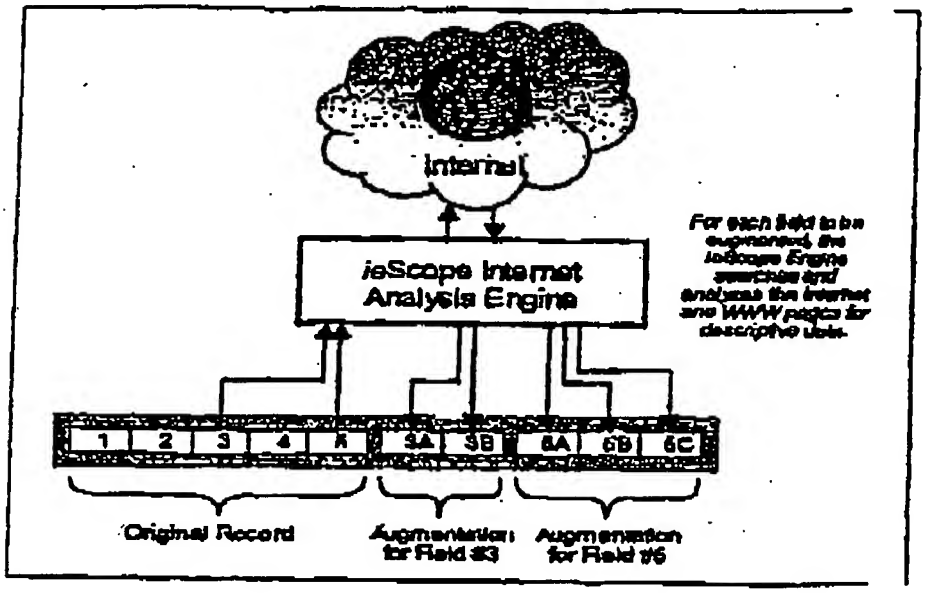
This document describes the core Internet mining technology behind *i*eScope, some of the general applications that can take advantage of this technology, and some of the industries to which this technology can be applied.

Contents

How It Works

At the highest level, *le*Scope works by taking transaction and record data from a large database, and searching the Internet for detailed descriptions and categorizations of that data. The objective is to utilize the Internet as the largest, most complete and most up-to-date description of the world. The descriptive data that is returned from the web searches is then added to the original data to provide organizations with a richer internal data source for various applications.

A general functional diagram of *le*Scope is shown below.



In this diagram, the original data record is made up of five fields. Two of these fields are to be augmented by information derived from the Internet and World Wide Web pages. *le*Scope takes the values in these two fields, and using the specific technologies within the Analysis Engine, searches the Internet and then condenses the results obtained about these specific values. After this analysis is complete, *le*Scope produces descriptive values for each of the fields. In this case, the system has been configured to produce two descriptive terms for Field #5 and three descriptive terms for Field #6. Examples of how this system may be applied to real-world data environments are listed in a later section.

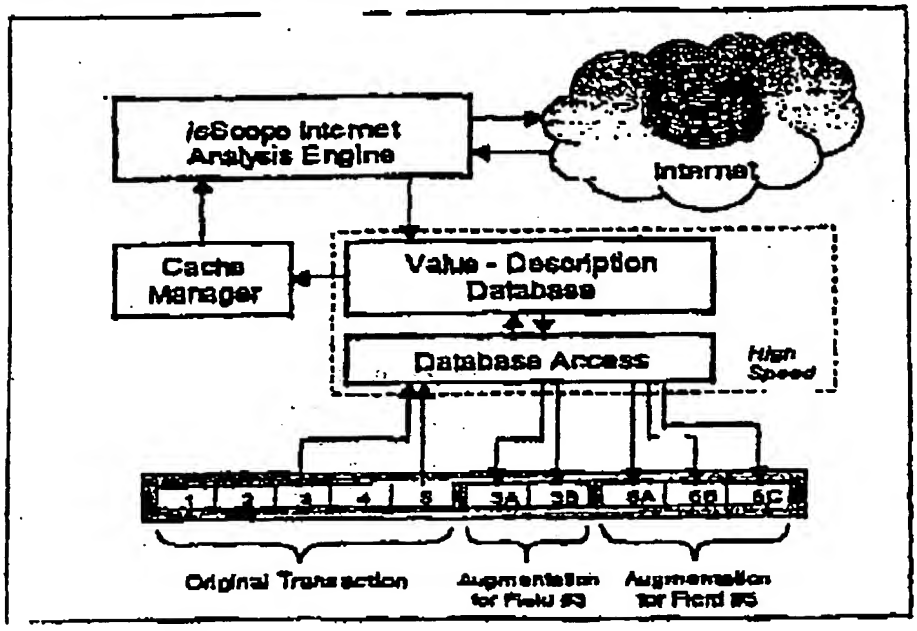
Smart Caching Technology

In its direct form, the system shown above will generally be too slow for modern large-scale data systems. Modern transaction processing systems are required to process in the order of 100 to 1000 transactions per second, making external access to a comparatively slow resource such as the Internet an impractical endeavor. This can be easily remedied by an intelligent caching system.

Data fields presented to *ieScope* for analysis will generally range in size from 1000 distinct values up into the millions of distinct values. Below about 1000, the field values are most likely to have explicit operational meaning, and the added value of *ieScope* will not be as valuable. The cardinality of data sets can be used to establish an intelligent caching system.

In addition to the cardinality issue, most descriptions that *ieScope* creates will be for entities that may change week-to-week or month-to-month, but are unlikely to be changing minute by minute. As such it will be appropriate to cache the descriptions for each value of a field, and update them only as required. This update can occur on a regular basis or upon actively detecting changes in the Internet resources from which the analysis has been done. In this manner, very high-speed transaction augmentation can occur, with periodic updates to the cache to keep the descriptions current.

The diagram below shows a high level configuration for the caching system.



In the above diagram, the Value-Description Database contains the cached descriptions for each field value. When an unknown value is encountered, this results in a cache-miss which is detected by the cache manager and passed to the *ieScope* Engine. The *ieScope* Engine executes the various Internet searches and other tasks, and creates the description. This is then put into the cache as a new value-description record.

The cache manager also utilizes the *ieScope* Engine to update the descriptions in the cache as required. This can be triggered on a time schedule, by detecting changes in the Internet source from which each given description was created, and/or by other external signals.

Because the database is relatively simple in structure (although potentially large in size), high transaction speeds will be easily achieved. In this manner this form of intelligent cache has enabled high-speed access while not compromising the quality of the core data augmentation process.

Example Applications of *IeScope*

This section describes a number of example applications for the *IeScope* technology. These are a subset of the full range of areas in which *IeScope* will be able to provide real value.

Payment Card Transaction Augmentation

Payment Card transactions contain only a rudimentary description of the actual merchant / product or service that was purchased. The key indicator typically used is the Standard Industry Code field (often referred to as the "SIC code") which is a four digit integer numeric field. Unfortunately the SIC code is not always accurately mapped, and has a wide variation in resolution. Some SIC codes identify individual companies, whilst others collapse an entire industry into a single code.

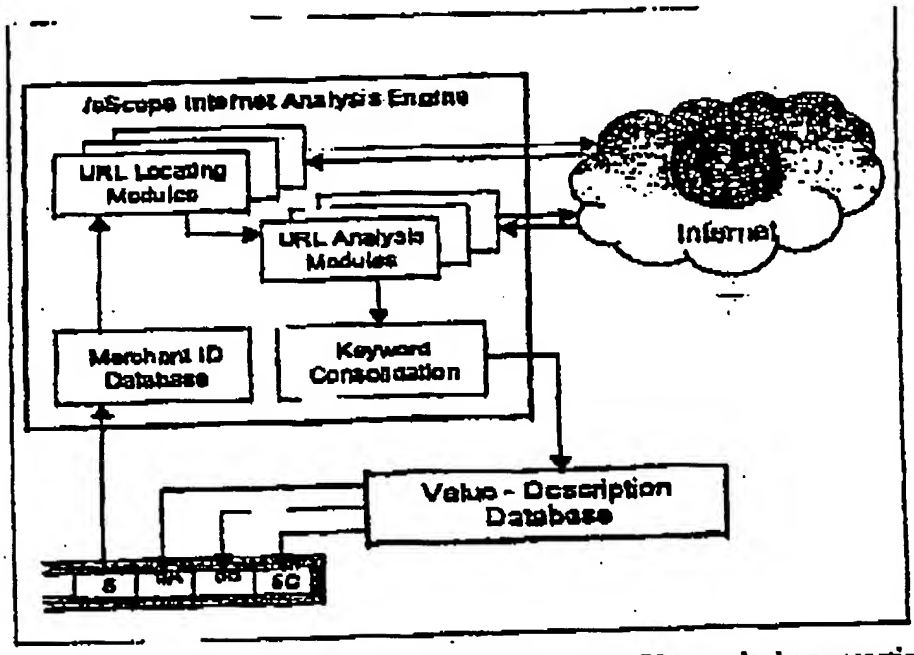
The other key descriptive field in the transaction is the merchant name. This is a text string that typically contains some form of the merchant's name. This field is usually used for statement generation purposes, and as such is often compressed and shortened. From an analytic viewpoint, even accurately knowing the company's name is of little value as translating the name into a business or product category is difficult.

This is where *IeScope* can add unique value. By taking the merchant name as the field to be augmented, *IeScope* can go onto the web and create a description for that merchant in terms of known English keywords. For example, a transaction at "Frank BMW" would return keywords such as "automobile" and "luxury" and "quality". A transaction at "Hotel Del Coronado" would return "resort", "historical", "Victorian" and "sea side".

The Keywords extracted are thus a reflection of the marketing pitch being made by the merchant. When one considers that it is typically this very pitch which has prompted the consumer to patronize the merchant, one starts to understand that such a representation of the consumer, would it be possible, becomes a powerful proxy for their purchasing predisposition based on the real world using a nomenclature which is easily understandable, namely the English language.

These keywords, in appropriate symbolic form, would then be added to the individual transaction record. For the first time, analysts and marketers would be able to search for and cluster their Payment card customers based on English descriptions of the products and services they purchase. A more detailed description of an analytic marketing tool based on *IeScope* is presented later.

There are various domain specific modules required to build *IeScope* for the Payment card industry. These are captured in the diagram below:



The Merchant ID Database holds the full set of merchant Ids seen in the transactions. A set of URL Locating Modules scans various Internet resources to locate appropriate reference URLs for each merchant Id. Once located, these URLs are passed into the URL Analysis Modules, which are specifically designed to extract the company and/or service descriptive characteristics from each of the different Internet resources.

The URL Locating Modules include the following:

- Direct access to company home pages based on a name search. This may be augmented through a lookup in the RealNames ™ database and/or company listings online.
- Searches of financial websites for company descriptions.
- Searches of industry web pages, local-listing web pages and other geographic specific resource sites. There are a large number of local listing pages for each city and/or state on the Internet today.

Each URL module may report multiple URLs that contain appropriate descriptions. The most suitable (as determined by the various modules) will be selected as being the definitive set for that merchant or service.

The URL Analysis Modules takes each of the resources provided by the URL Locating Modules and condense them appropriately to construct the final set of

descriptive keywords. These modules are designed around specific Internet resources, including the following:

- Company "about" and "history" pages.
- Company quarterly / annual reports and letters to shareholder.
- Financial web-site corporate profiles.
- Aggregate web-site spider analysis.
- Standard Industry Classifications (SIC) code databases.
- Search engine summarized descriptions and categorizations (where available). This includes descriptions from RealNames™.

In an offline environment, full population keyword analysis is done to ensure that the keywords being extracted are sufficient to describe the merchants and service providers. Because the descriptions are human readable, random quality assurance is possible to determine the effectiveness of each of the modules described above. Because these modules are reasonably heuristic, improvements can be made iteratively by analyzing and comparing the performance of each module on its target subset of the data.

Retail Product Description

Retailers will be able to use the same technology to categorize products into a more regularized description. Product names can be indexed against the Internet through manufacturer web pages, product review sites, online merchandisers and so on. The customer's retail card purchase transaction record is then used as the index. As in the credit card example above, all data in the consumer's transaction history over the desired time period is then augmented by the Internet derived data. The result, a powerful description of the consumer.

The applications that could utilize this data include general marketing and cross-sell opportunities, as well as tracking individual customer product preferences at a more descriptive level.

Telecommunications Call Destination Description

Telephone call destination numbers can be indexed to company names and from there indexed to descriptions of those companies. This would require an additional telecommunications specific module to convert from the destination telephone number to a company or service provider name. Personal phone calls to residential numbers would be utilized in this system to the extent that the system can match personal phone numbers to personal web content.

Credit Bureau Augmentation

There are a small number of fields that could be augmented in a credit bureau. For example, where the employer's name is known or where loan collateral is described on the credit report this information can be indexed against the web to more fully describe it. This may add useful data for credit approval and general credit application processing.

Collections

In the collections environment, *ie*Scope can be used to describe the transaction history of the individual at any point in time up to and including the time that they went delinquent. The additional information would be valuable to determining the specific actions to be taken on each collections case or whether the debt should be managed in house or sold to a third party.

Investment Data

The securities an individual invests in are an indication of the type of person they are, where their interests lie and their propensity to assume risk. *ie*Scope can extract descriptions of companies in the same manner described above and use it to describe the individual in question. The recent federal ruling that allows financial institutions to offer securities to their account holders would allow us to combine the *ie*Scope investment keyword descriptions with the payment card keywords described above.

In a large bank that offers cross-industry services, this data may be useful in cross-sell opportunities.

Affinity Program / Cross Sell Targeting

For a financial institution, airline or other large company that runs and manages an affinity program, *ie*Scope can take advantage of the transaction and account data to better target product offers to individuals. Offers can be described in English language terms and then used to match against the individuals' keyword descriptions.

In a large financial institution, financial products could be described by keywords, which could then be matched against customer keyword descriptions and used to target one-to-one offers. The deployment of the offer would be via the institutions Internet web site.

Because an individuals Keyword profile is changing in real-time, the opportunity to target the offers at the appropriate time for any individual consumer is greatly enhanced.

News and Personalization Products

Whenever an institution wishes to personalize offerings such as a portal web-site, news services, information services and so on, they can use *i*eScope to assist in describing the interests and motivations of their customers, to the extent that transaction and account data is available for this purpose.

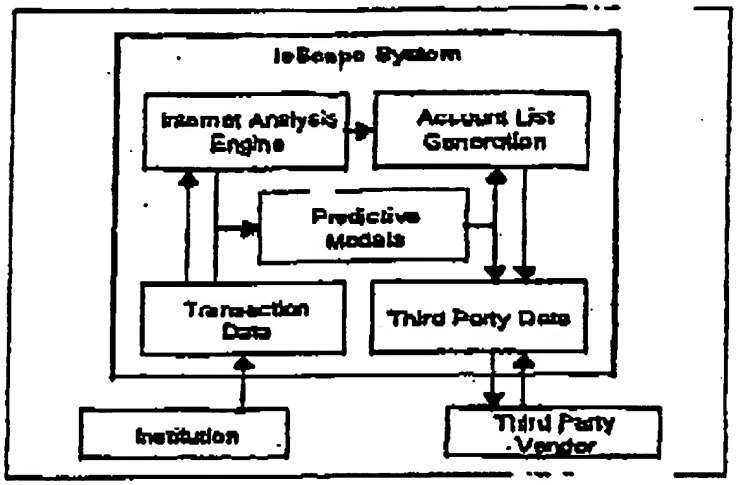
Third Party List Specification and Distribution

Some institutions provide lists of customers for third party companies that sell products to those people. These lists may be used by the vendor in a number of ways, such as through telemarketing or direct mail. Typical products may include insurance, financial services, travel services, as well as more general merchandise and services. Due to privacy concerns, the providing institution restricts the types of products to those that it feels are suitable products for it to be associated with.

*i*eScope provides a unique ability for either the third party vendor or the institution owning the data to specify and filter these lists according to characteristics of individuals. In this scenario, the vendor may provide the search terms and other qualifying characteristics, and would be given back the specified list of customers. Clearly the actual fields that a vendor would have access to would be restricted from that which the bank would have access to. This system provides the vendor with a convenient form of specifying and accessing lists.

*i*eScope also provides the ability for enhanced predictive modeling in cases where vendor offer response data is available. In this scenario, response information is provided back to us where predictive models are built. These predictive models utilize the *i*eScope data as key predictive inputs. Once built, the predictive models can be used to extract subsequent lists for the vendor of even higher accuracy. The addition of this service then permits the vendor to have a wide choice of options for specifying lists.

The diagram below shows a possible system setup including the granting of restricted access to the system from third party vendors.



The diagram above shows a high level flow of data for a scenario in which a third party vendor wishes to utilize the *IeScope* search and list building system. Access would generally be granted under the control of the institution providing the data. The components in this system have the following roles:

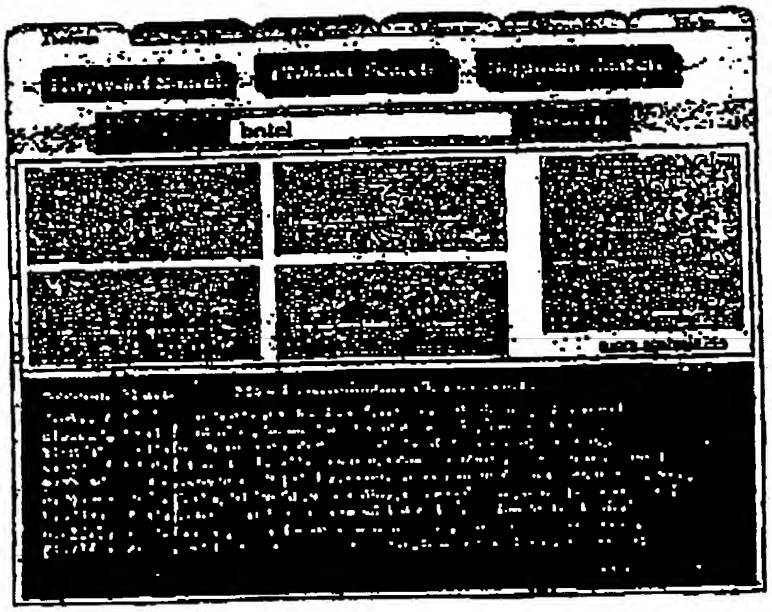
- **Institution:** Provides the data (account, transaction, etc).
- **Transaction Data:** Data warehouse maintained within the *IeScope* system.
- **Internet Analysis Engine:** *IeScope* internet analysis engine that searches the internet, analyzes found pages and augments the transaction data.
- **Account List Generation:** Modified form of the keyword based search engine and list builder that limits data access to only those fields and/or data and transactions that the providing institution wishes to disclose.
- **Third Party Data:** The third party builds their lists. In the case of targeted marketing or other predictive problems, these lists may be augmented by response or other outcome data based on the third party vendor's actions carried out on the lists. In the case of direct marketing, response data to an earlier direct marketing campaign would be provided by the third party vendor.
- **Predictive Models:** Incorporating the data from *IeScope*, transactional data and the third party vendor response data, predictive models can be built. These can then be subsequently used to filter lists for even more refined list creation. These predictive models could also be used to predict credit risk, attrition risk, bankruptcy, fraud and other standard indicators. These latter applications would require some additional infrastructure, but utilize the *IeScope* data very effectively.

IeScope Powered Applications

In the previous section, IeScope was presented as a data augmentation system for a diverse range of industries. Of key importance here is the fact that IeScope is automatically describing the transaction data in terms of well-defined English keywords. This is both a powerful and very intuitive form of data to work with. This section illustrates this with applications that can extract operational and analytic value from this data in very easy to manage ways.

Segmentation by Basic Search

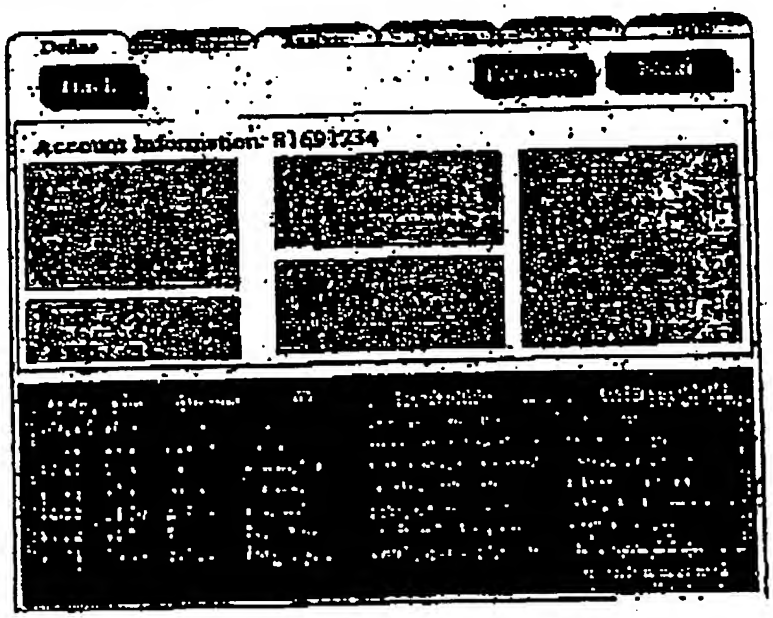
A portfolio of accounts can be segmented based on keyword description, in very much the same way that web pages can be located using a keyword based search engine. In a Payment card example, a marketing manager may wish to market a product to all people who stay at hotels. In this simple case, they simply enter the keyword "hotel" into the search window and a list of individuals appears who have visited hotels:



In the screen above, the subset of data has been defined as simply those accounts that most match to the keyword "hotel". Options on this include the ability to generalize "hotel" to higher-level categories and to make web-page associations between keywords and other keywords to enhance the matching.

The analysis presented on this page is an example only. The example accounts can be rank ordered by degree of match to the keywords. The user can select an account to view the transaction history and the description of this individual. For the purposes of marketing, this would be useful for doing a small sampling of the customers and determining their suitability to the program being designed.

An example of the individual account screen is shown below:



In this screen is shown basic account information, the top keywords that describe this account, the top discriminator keywords for that account (keywords that are generally rare but strong in this individual account), and their transaction history. Each transaction is shown with the accompanying descriptive terms for that transaction and a reference URL. The purpose of showing the reference URL is to enable the user to further explore the individual characteristics of the customers that fall into this particular list. Other analysis will be applied later to group the reference URL and other information at the list level.

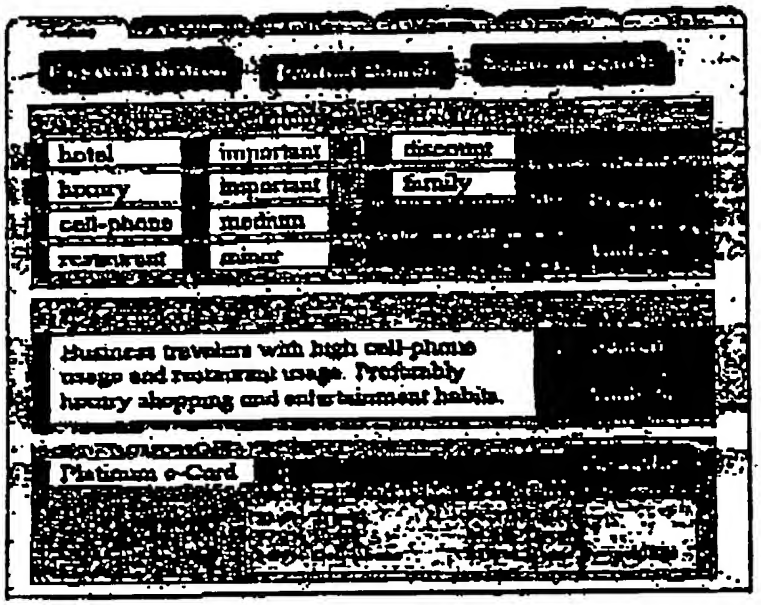
Segmentation by Advanced Search

The lists of customers produced by the basic keyword search will be interesting to peruse, and possibly to utilize in a marketing campaign. However, as with web-based searches, there is significant additional value to be obtained from advanced search capabiliti

For example, these include:

- Searching by multiple keywords with weightings on different keywords.
- Searching by multiple keywords and associated keywords. The lists of associated keywords can be derived from web-page co-occurrence or other sources.
- Exclusion rules to remove customers with strong affinity in certain keywords and categories.
- The ability to enter a set of keywords and, from the list of matching customers, derive a second set of keywords that illustrate "like behavior" for that list.
- The ability to enter a product description in English language form and use this as the set of search terms to build a list.

The diagram below shows some examples of this more advanced search.



Segmentation by Clustering

Segmentation can also be done based on clustering. In this approach, similar accounts, transactions and/or merchants can be grouped together based on similarities in their underlying characteristics. The key advantage of *IcScope* in this realm is that there is additional descriptive data on which the clustering can be done. This will be especially valuable in situations where typically only numeric data exists.

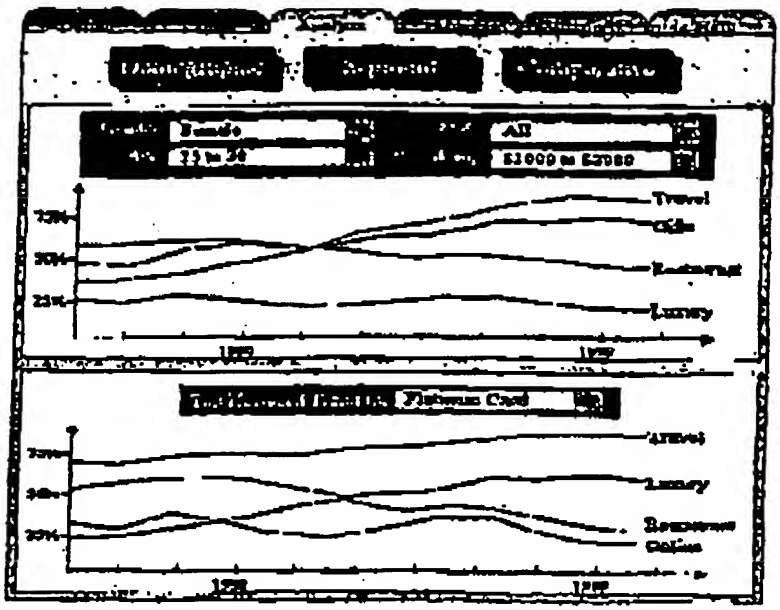
There are many segmentation algorithms in existence. Most of these will be able to utilize the added data provided by *IcScope*.

Trend Analysis

IcScope describes transactions and accounts in real-world terms. While this is very valuable, it is also important to look at how accounts, products and segments are changing over time. Trend analysis techniques can be applied to the *IcScope* enhanced descriptions to provide trending information for a variety of data categories including:

- Account level trends of keyword importance over time.
- Segment level trends by monitoring the importance of keywords within a defined segment over time.
- Segment trends can be defined on an individual product to look at how the underlying population that uses that product is changing over time.
- General customer demographic trends can be identified by looking at segment trends in segments defined by age, gender, geographic region and so on.

This type of trend analysis can be represented in a number of ways, and some examples are shown in the diagram below:



Cross-sell Analysis

A product can be described in terms of its current users and in terms of the desired market for the product. InScope can facilitate cross-sell opportunities by examining the key characteristics of individuals who are currently using the product and then finding similar individuals in whom the product would be applicable. Because this is based on latent/visual keyword matches it gives better control at each stage of the cross-sell process.

A cross-sell process could involve the following steps:

1. **Determine the key characteristics of the target product:** This can be done either by entering desired keywords for the target product, or by using the system to determine the set of most discriminating keywords that describe the current users of the product (versus the overall population), or a combination of both.
2. **Build A Segmentation:** Using the new description, build a segmentation of individuals who currently do not have the product but most closely match the description of the product.
3. **Refine the Segmentation:** Refine the segmentation against standard criteria such as demographic filters (age, state, zip, gender, etc) and against

operational characteristics (current balances, profitability, revolving rate, etc).

Modeling Applications

isScope adds valuable data to transaction and account records. This data can be used in traditional predictive modeling to build models such as:

- Risk models (bankruptcy, attrition).
- Profitability.
- Fraud (assist in detecting lost and stolen and counterfeit cards).
- Customer service.
- Third party vendor offers.

These models can utilize standard numerical and symbolic data, and there are a wide variety of techniques available to do this. The value of *isScope* in this domain is the addition of valuable descriptive data.

Database Query Interface

The *isScope* technology can be used to create a natural language interface to traditional databases. In this, data fields from the database records are mapped into categories (by an expert familiar with the database). *isScope* creates descriptions of the individual entries in keyword or key-phrase terms from the data sources described in this document, and potentially other data sources that are specific to the data field being described.

The search, clustering and modeling technology can then be used to query the descriptions on the database, create clusters, build predictive models and so on. This may provide enhanced access to data that would otherwise require expert understanding to work with. Examples of this idea are:

- Converting job titles to descriptions of job skills and desired employee characteristics. In this way, people could search their employee database for "analytic" and come up with engineers and accountants ranked higher than marketing or sales.
- Converting product names or codes into product descriptions. This would allow people to search for a product or group of products when they do not know the particular industry jargon. This example highlights the ability of *isScope* to translate information – in this case from product names into common-use English terms.

- Mapping scientific names into descriptions. For example, non-experts could search a specially designed taxonomic database (such as trees, animals, seeds, diseases, chemicals and so on) using non-expert language.

Additional Analysis

Advertising Analysis

The data that *leScope* collects is, to some degree, a first-order cross-reference to the Internet. Higher order analysis is also possible, for example:

- In analyzing a merchant's website certain keywords may not be well explained from the website alone. Additional related keywords can be extracted from other related websites. For example, a BMW service center may not describe BMWs very well. The keyword BMW may best be described from another site or set of sites. This second-order cross-reference can be especially valuable to enhance small merchant's web sites where it is not clear what the quality or marketing message is behind the products.
- Additional correlative value can be extracted by characterizing where a product is advertised. Luxury goods are more likely to be advertised on financial websites, whereas more average priced goods are more likely to be advertised on sports or entertainment sites. Spidering the web and looking at the relationships between products, where they are advertised and with what other products are they advertised will provide useful correlative information. This is only possible for Internet advertising where the information is in an easy-to-access digital form that can be processed automatically from a spidering system.
- Additional data can be extracted by looking at which web-pages link to the merchant reference URLs, and which pages they link to. This will identify partner programs and other related merchants.

Additionally, the *leScope* technology could easily be adopted to allow web-marketing companies like DoubleClick and AdForce to better understand consumer web behavior. Currently, these companies primarily target users by data such as time of day, what area a person lives or works in and perhaps what company they work for, if they are in school, and what Web site they are looking at. *leScope* would add real-world keyword descriptions of the web sites visited by an individual in the same manner it did for Payment Card transactions. The result would be a richer description of the online consumer

Population Description

Because it works in the domain of English language keywords and expressions, *leScope* can provide real-world descriptions of populations and segments of populations. This may be of value in determining the key characteristics of different product users for the purposes of driving advertising and marketing content. In this way, *leScope* can actually provide the content creators with descriptive information that may help them better understand their audiences, and hence better design advertising material.

Other Data Sources

In this description of *isScope*, the Internet is described as the key source for complete and up-to-date information about the world. This is certainly an appealing and attractive source of data, and its importance is growing every day. However other specifically accessed data sources may provide benefit, such as:

- Credit bureau data, especially relating to specific loans granted or applied for.
- Electronic encyclopaedia, dictionary and thesaurus data. Although this is not at all dynamic, it should provide some useful generalization ability to *isScope*.
- News services not directly available over the Internet.
- Electronic forms of the yellow and white pages, census data and general contemporary literature. If, for commercial reasons, this data is not available on the Internet, we may be able to purchase it for the building of the *isScope* knowledge base.
- Internet transactions. We can describe an individual by the keywords associated with general Internet browsing. The specific keywords may be different, but we can build up a profile based on spidering the websites and pages that they visit.